HHS Public Access
Author manuscript
Epidemiology. Author manuscript; available in PMC 2016 January 29.

Published in final edited form as:
Epidemiology. 2014 November ; 25(6): 913–916. doi:10.1097/EDE.0000000000000181.

Bias in calculation of attributable fractions using relative risks from non-smokers only

Katherine M. Flegal
National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville MD, USA

Abstract

Studies of weight and mortality sometimes state that the mortality relative risks for obesity from non-smokers are valid estimates of the relative risks for obesity in both smokers and non-smokers. Extending this idea, several influential articles have used relative risks for obesity from non-smokers and attributable fraction methods for unadjusted risks to estimate attributable fractions of deaths in the entire population (smokers and non-smokers combined). However, stratification by smoking is a form of adjustment for confounding. Simplified examples show that the use of relative risks from only one stratum to estimate attributable fractions, without incorporating data on the stratification variable, gives incorrect results for the entire population. Even if the mortality relative risks for obesity from non-smokers are indeed valid in both smokers and non-smokers, these relative risks nonetheless need to be treated as adjusted relative risks for the purpose of calculating attributable fractions for the whole sample.

In epidemiologic studies of obesity as a risk factor for mortality, it is sometimes recommended to calculate mortality relative risks for obesity from a sample of only non-smokers because it is felt to be difficult to adjust statistically for smoking.[1, 2] For example, Berrington de Gonzalez et al. [1 p. 2217] state that "Stratification or exclusion rather than adjustment is necessary because smoking is so strongly related to obesity and mortality." An extension of this is the idea that the mortality relative risks for obesity from non-smokers represent more valid relative risks for obesity in both smokers and non-smokers and thus should be used to calculate population attributable fractions (PAFs) for obesity in the whole population, including both smokers and non-smokers. For example, Calle et al [3 p. 1634] state "The estimates based on relative risks among men and women who never smoked … do not describe the fraction of deaths attributable to overweight and obesity among this population only. Rather, they are estimates of the fraction of deaths attributable to overweight and obesity in the total U.S. population, on the assumption that the relative risks among those who never smoked offer the most valid estimates of the true effect of overweight and obesity on mortality from cancer."

Corresponding author: Katherine Flegal, National Center for Health Statistics, Centers for Disease Control and Prevention, 3311 Toledo Rd, Room 4336, Hyattsville MD 20782, Tel; 301-458-4215, Fax: 301-458-4029, kmf2@cdc.gov.

Required disclaimer: The findings and conclusions in this report are those of the author and not necessarily the official views of the Centers for Disease Control and Prevention

Calle et al [3] and others [4, 5] have used relative risks for overweight and obesity from never-smokers to calculate attributable fractions from all deaths occurring in a target population that lacks data on smoking status among decedents. Here I present some simplified examples to show the potential errors introduced by this procedure. I compare the results from using two different computational formulas described by Rockhill et al [6] for PAF calculations from generated data sets. Following the notation shown in Table 1 of Rockhill et al, Formula 1 is $p_e*(RR-1)/(p_e*(RR-1)+1)$ where p_e is the proportion of the population exposed to the factor (in this case, to overweight) and RR is the relative risk of the outcome (in this case mortality) associated with overweight. Formula 2 is $pd*(RR-1)/RR$, where pd is the proportion of cases exposed to the risk factor (in this case, the proportion of the decedents who were overweight) and RR is the relative risk of mortality associated with overweight. This is a formula appropriate for use with adjusted relative risks when confounding exists. [6, 7]

These computational formulas are valid for calculations, but, like other computational formulas, can give rise to misunderstandings and be used inappropriately because they are not definitional formulas that describe the underlying relationships.[8, 9] Perhaps because these formulas do not describe the underlying relationships, attributable fractions are frequently calculated incorrectly.[6] When there is no confounding, Formula 1 and Formula 2 are algebraically identical to each other. However when relative risks are adjusted for confounding, these formulas are not equivalent, and Formula 2 should be used instead of Formula 1. Rockhill et al. [6] stated that “probably the most common error” was to calculate attributable fractions with adjusted relative risks in a formula, such as Formula 1, that is appropriate only for unadjusted relative risks, an approach that has continued to be used.[3–5, 10–15] Because stratification is a form of adjustment for confounding, [16 p. 176ff] it needs to be taken into account when calculating attributable fractions. If data on the numbers of deaths within each stratum were available, attributable fractions could be calculated within strata and summed over the population by using the weighted sum method.[7, 17] However, the required information (e.g. the proportion of decedents who are smokers) is often not available.

Small example data sets for illustrative purposes were generated using the approach outlined by Darrow and Steenland.[18] Table 1 shows a generated data set in which the risk factor is overweight, the stratification variable is smoking status (smoker or non-smoker) and the outcome is mortality. In this example, smoking is a confounding factor, since it is associated with a lower prevalence of overweight and a higher risk of mortality. The attributable fraction is calculated as the sum of the category-specific differences between observed and expected, divided by the sum of the observed numbers. There are 162 deaths in this example, but if there were no excess risk associated with overweight, there would be 150 deaths. The difference, 162–150, represents excess mortality associated with overweight. The “true” attributable fraction is thus 12/162=0.074 or 7.4% of all deaths.

The generated data sets varied only in mortality relative risks for overweight and for smoking and were otherwise identical. The results from using Formulas 1 and 2 for calculating PAF for the whole sample from generated data sets with different combinations of relative risks for smoking are shown in Table 2. All examples have the same mortality relative risks for overweight in both smoking strata and thus correspond to the assumption that the relative risks for overweight from non-smokers are the correct relative risks for overweight in smokers as well. The mortality relative risks for smoking vary across examples, but in all examples the relative risks for smoking are the same in both weight strata.

Example 1 in Table 2 is the same as the example in Table 1. For this example, as shown in Table 2, Formula 1 yields a PAF of 9.1% for the sample and Formula 2 yields the correct PAF of 7.4%. However, the use of Formula 2 requires information on the proportion of deaths occurring among non-smokers-- information that is often not available. Because the mortality relative risks for overweight are adjusted relative risks (adjusted by stratification), Formula 1 does not give the correct result when it is applied to deaths in the whole population without taking the stratification into account. As shown in this example, even if the relative risk for overweight calculated from only non-smokers is the correct relative risk for overweight in both smokers and non-smokers, nonetheless the estimate from using Formula 1 will generally be biased if the stratification by smoking is not considered. This bias arises because the relative risk for overweight is adjusted via stratification for confounding by smoking, but the attributable fraction method used does not consider the stratification and does not account properly for the effect of smoking on mortality. We have called this elsewhere[17] the “partially adjusted” method of calculating attributable fractions, because the relative risks are adjusted for the confounder but the attributable fraction calculation is not adjusted for the effects of the confounder on mortality.

As may be seen, the true PAF values for overweight vary with changes in the mortality relative risk for smoking, because these changes affect the degree of confounding by smoking. Within smoking strata, there is no confounding by smoking. Thus Formula 1 and Formula 2 both give the correct PAF values for non-smokers and for smokers, considered separately. However, the PAF estimates for overweight in the whole sample using Formula 1 are invariant to changes in the mortality relative risk for smoking, showing again that this approach does not appropriately account for confounding by smoking.

In examples 6–11 in Table 3, the relative risk of mortality associated with overweight differs by smoking strata. In this case, neither Formula 1 nor Formula 2 gives correct estimates for PAF when used with relative risks for overweight from non-smokers only. The magnitude of the error varies according to the mortality relative risks for smoking as well as with the mortality relative risks for overweight.

In these examples, the bias tends to be upward, but the direction of bias varies. Attributable fractions are complex functions that depend on the prevalence of the exposure within confounder strata, the prevalence of the confounder within exposure strata and the relative risks within each exposure-confounder subgroup. As a result, the magnitude and the direction of expected bias are not easily generalized. Darrow and Steenland [18, p. 53] found that “Bias in the AF [attributable fraction] increases as the magnitude of the confounding increases, and is dependent on the prevalence of exposure in the total population, with bias greatest at the lowest prevalence of exposure. Bias in the AF is also higher when the exposure-disease association is weaker.”

The relative risks of mortality for overweight from non-smokers are adjusted, by stratification, for confounding by smoking. As shown in the simple examples provided here, attributable fractions of deaths in the entire sample cannot appropriately be calculated using these relative risks in a formula for unadjusted risk and then applying them to the entire sample while ignoring the stratification variable. Examples of this error can be seen in the papers by Allison et al, [4], Calle et al, [3] and Mokdad et al. [5]. The degree of bias from this approach is affected by the strength of the confounding by smoking. Even if the mortality relative risks for obesity from non-smokers are indeed valid relative risks for obesity in both smokers and non-smokers, these relative risks nonetheless need to be treated as adjusted relative risks for the purpose of calculating attributable fractions for the whole sample.

Acknowledgments

Financial support: No external financial support

References

1. Berrington de Gonzalez A, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis RJ, et al. Body-mass index and mortality among 1.46 million white adults. The New England journal of medicine. 2010; 363(23):2211–9. [PubMed: 21121834]
2. Stampfer M. Weight loss and mortality: what does the evidence show? PLoSMed. 2005; 2(6):e181.
3. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. NEnglJMed. 2003; 348(17):1625–38.
4. Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB. Annual deaths attributable to obesity in the United States. JAMA. 1999; 282(16):1530–8. [PubMed: 10546692]
5. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA. 2004; 291(10):1238–45. [PubMed: 15010446]
6. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. AmJPublic Health. 1998; 88(1):15–9.
7. Benichou J. A review of adjusted estimators of attributable risk. StatMethods MedRes. 2001; 10(3): 195–216.
8. Guttmannova K, Shields AL, Caruso JC. Promoting conceptual understanding of statistics: Definitional versus computational formulas. Teaching of Psychology. 2005; 32:251–3.
9. Muttart DM. The obsolescence of computational formulae. Teaching Statistics. 2009; 31(1):12–4.
10. Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJ, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoSMed. 2009; 6(4):e1000058.
11. Ezzati M, Hoorn SV, Rodgers A, Lopez AD, Mathers CD, Murray CJ. Estimates of global and regional potential health gains from reducing multiple major risk factors. Lancet. 2003; 362(9380): 271–80. [PubMed: 12892956]
12. Flint AJ, Hu FB, Glynn RJ, Caspard H, Manson JE, Willett WC, et al. Excess weight and the risk of incident coronary heart disease among men and women. Obesity (Silver Spring, Md). 2010; 18(2):377–83.
13. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980–2000. NEnglJMed. 2007; 356(23):2388–98.
14. Forman JP, Stampfer MJ, Curhan GC. Diet and lifestyle risk factors associated with incident hypertension in women. JAMA. 2009; 302(4):401–11. [PubMed: 19622819]
15. Preston SH, Stokes A. Contribution of obesity to international differences in life expectancy. Am J Public Health. 2011; 101(11):2137–43. [PubMed: 21940912]
16. Rothman, KJ. Epidemiology: An Introduction. 2. Oxford University Press; 2012.

17. Flegal KM, Graubard BI, Williamson DF. Methods of calculating deaths attributable to obesity. AmJEpidemiol. 2004; 160(4):331–8.

18. Darrow LA, Steenland NK. Confounding and bias in the attributable fraction. Epidemiology. 2011; 22(1):53–8. [PubMed: 20975564]

Table 1

Schematic example[a] of calculation of excess mortality associated with overweight among smokers and non-smokers when the relative risk associated with overweight is the same in smokers as in non-smokers

Smoking Status	Weight status	No.	Observed no. deaths	Expected no. deaths if normal weight	Excess no. deaths associated with overweight (Observed minus expected)
Non-smokers	Normal	300	30	30	0
	Overweight	450	54	45	9
Smokers	Normal	200	60	60	0
	Overweight	50	18	15	3
Total		1000	162	150	12

[a] Generated based on a 10% risk of mortality among normal weight non-smokers, a prevalence of overweight of 50%, a prevalence of smoking of 40% among normal weight and 10% among overweight, a relative risk of 1.2 for overweight and of 3 for smoking

Table 2

True and calculated values of overweight-associated population attributable fractions (PAFs)[a] for generated data sets with varying relative risks for smoking when relative risks for overweight are identical across smoking strata

	Example 1	Example 2	Example 3	Example 4	Example 5
Relative risks of mortality for:					
Overweight relative to normal weight among non-smokers	1.2	1.2	1.2	1.2	1.2
Overweight relative to normal weight among smokers	1.2	1.2	1.2	1.2	1.2
Smoking relative to non-smoking among those of normal weight	3	4	2	3	3
Smoking relative to non-smoking among overweight	3	4	2	5	2
PAF values for whole sample:					
True PAF	7.4	6.9	8.1	8.0	7.1
PAF1	9.1	9.1	9.1	9.1	9.1
PAF2	7.4	6.9	8.1	8.0	7.1

[a] PAF 1 is calculated from Formula 1 using the RR for non-smokers.

PAF 2 is calculated from Formula 2 using the RR for non-smokers.

Table 3

True and calculated values of overweight-associated population attributable fractions[a] for generated data sets with varying relative risks for smoking and relative risks for overweight that differ by smoking strata

	Example 6	Example 7	Example 8	Example 9	Example 10	Example 11
Relative risks of mortality for:						
Overweight relative to normal weight among non-smokers	1.2	1.2	1.2	1.2	1.2	1.2
Overweight relative to normal weight among smokers	1.1	1.1	1.1	1.3	1.3	1.3
Smoking relative to non-smoking among those of normal weight	3	3	3	3	3	3
Smoking relative to non-smoking among overweight	3	5	2	3	5	2
PAF values:						
True PAF	6.5	6.7	6.5	8.3	9.3	7.6
PAF1	9.1	9.1	9.1	9.1	9.1	9.1
PAF2	7.3	7.9	7.0	7.5	8.2	7.1

[a] PAF 1 is calculated from Formula 1 using the RR for non-smokers.

PAF 2 is calculated from Formula 2 using the RR for non-smokers.